FEDERAL RESERVE BANK *of* ATLANTA

Confronting Model Misspecification in Macroeconomics

Daniel F. Waggoner and Tao Zha

Working Paper 2010-18
December 2010

WORKING PAPER SERIES

FEDERAL RESERVE BANK *of* ATLANTA **WORKING PAPER SERIES**

Confronting Model Misspecification in Macroeconomics

Daniel F. Waggoner and Tao Zha

Working Paper 2010-18
December 2010

Abstract: We confront model misspecifications in macroeconomics by proposing an analytic framework for merging multiple models. This framework allows us to address uncertainty about models and parameters simultaneously and trace out the historical periods in which one model dominates other models. We apply the framework to a richly parameterized dynamic stochastic general equilibrium (DSGE) model and a corresponding Bayesian vector autoregressive model. The merged model, fitting the data better than both individual models, substantially alters economic inferences about the DSGE parameters and the implied impulse responses.

JEL classification: C52, E2, E4

Key words: merged model, misspecification, state-dependent weights, model uncertainty, parameter uncertainty, impulse responses, policy analysis

The authors thank Frank Diebold, John Geweke, Frank Schorfheide, and Chris Sims for helpful discussions. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Daniel F. Waggoner, Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street, N.E., Atlanta, GA 30309-4470, 404-498-8278, daniel.f.waggoner@atl.frb.org, or Tao Zha, Federal Reserve Bank of Atlanta and Emory University, Research Department, 1000 Peachtree Street, N.E., Atlanta, GA 30309-4470, 404-498-8353, tzha@earthlink.net.

I. Introduction

A stochastic dynamic equilibrium, indexed by a *parameterized* model, is a likelihood function. Given the likelihood and the prior density of model parameters, one can simulate the posterior distribution and compute the marginal data density (MDD). The MDD is then used to measure how well the model is fit to the data.

Consider the situation in which there are multiple models on the table. The conventional procedure for model selection is to compare MDDs amongst individual models.[1] Since it is not uncommon that the MDD implied by one of the models is overwhelmingly higher than the MDDs implied by others, this procedure often ends up with the selection of one model at the exclusion of others. One primary example is that a linearized dynamic stochastic general equilibrium (DSGE) model such as Smets and Wouters (2007) can easily trump a standard Bayesian vector autoregression (BVAR) model.[2] The implication is that the BVAR can be simply replaced by the DSGE model for policy analysis.

Despite such overwhelming evidence presented by the posterior odds ratios in favor of one model, economists nonetheless continue to use both the DSGE and BVAR models in macroeconomic analysis. The tension between what the conventional procedure concludes and what actually transpires is a mere manifestation of increasing concerns about model misspecification by choosing a particular model (a particular likelihood) and categorically rejecting other models. Policymakers, as well as academic researchers, recognize that models are only *approximations* (Hansen and Sargent, 2001; Brock, Durlauf, and West, 2003; Sims, 2003). Indeed, they seldom rely on one single model even though this model fits better than other models according to the posterior odds criterion, because they know that

We confront model misspecification by proposing a Bayesian approach to merging multiple models. The merged model assigns *state-dependent* weights to predictive densities (conditional likelihoods) implied by different models so that the relative importance of each model changes across time. This new methodology, built on Geweke and Amisano (forthcoming), is motived by practical policy analysis dealing with situations where there are multiple competing models and each model explains (predicts)

[1]We implicitly assume that the prior weight is the same for all models. If the prior weight varies across models, we simply adjust the Bayes factors and calculate the posterior odds ratios.

[2]For sensitivity analysis, we also consider a case in Section VI.4, where the BVAR model trumps the DSGE model.

an observed outcome better than other models but only for certain episodes. An informal way for policy analysis is to employ a different model at a different time. Unlike the conventional model-averaging method, our Markov-switching approach not only assigns a weight of relative importance to each model but, more importantly, allows researchers to trace out the periods in which the data give the most weight to a particular model.

We apply our analytic framework to two widely used models: a richly parameterized DSGE model and a corresponding BVAR model. The MDD for the DSGE model is much higher than the BVAR model. The conventional Bayesian model-averaging method would imply that the BVAR model should receive nearly zero weight, a pathology discussed in Sims (2003). Our Bayesian approach overcomes this pathology. The merged model does not degenerate into the DSGE model or the BVAR model. To the contrary, our estimation indicates that the BVAR model dominates the DSGE model throughout two thirds of the history. The merged model, assigning nontrivial state-dependent weights to both models, fits the data considerably better than either the DSGE model or the BVAR model.

The rest of the literature has often treated the BVAR model as a benchmark to gauge how misspecified the DSGE model is. Our estimated results challenge this thinking because both the BVAR model and the DSGE model may be potentially misspecified. Rather than divorcing the data analysis from a particular model whose fit may not be as good, the estimation of our merged model indicates that both DSGE and BVAR models are operative but *at different times*.

Our methodology makes it econometrically implementable to establish the two-way communication between the theoretical DSGE model and the atheoretical BVAR model. We find that the posterior distributions of a number of key DSGE parameters change substantially when we incorporate the BVAR model in the merged model space. The error bands around impulse responses are predominately wider as the data imply more uncertainty about the DSGE model when the merged model is estimated. The relative importance of a structural shock in the DSGE model in explaining macroeconomic fluctuations is influenced heavily by the presence of the BVAR model. Thus, our approach integrates the two types of uncertainty, model uncertainty and parameter uncertainty, in one coherent framework.

II. Literature review

Our key assumption in this paper is that the true data generating process may not be among the models whose forecasts are combined. This insight appears in Diebold (1991), who argues that the standard Bayesian posterior-odds forecast averaging should be re-thought. Geweke and Amisano (forthcoming) propose a method of pooling the models by combining the predictive densities, which are consistent with out-of-sample forecasts. Although Geweke and Amisano (forthcoming) do not take a stand on the true data generating process, the log predictive score of pooled models tends to dominate the score of each individual model in the pool as the sample becomes large. This result is consistent with the extension of Geweke and Amisano (forthcoming)'s idea by Fisher and Waggoner (2010), who assume explicitly that the data generating process is a mixture of multiple models.

Our econometric methodology builds on these previous works. We show that state-dependent weights not only include Fisher and Waggoner (2010) as a special case but also gives a different interpretation about the relative importance of each model by estimating the probability that each model is chosen at time t.[3] Using the log predictive score, Geweke and Amisano (forthcoming) estimate the weights of models while taking the parameters in each model as given. Unlike Geweke and Amisano (forthcoming), we estimate the weights and the parameters of all the models jointly. One of our key findings is that the estimated parameters for the merged model are different from those when the models are estimated separately.

Del Negro and Schorfheide (2004) address potential BVAR misspecification by introducing the prior implied by a DSGE model into a BVAR model. We extend their idea by allowing for the two-way communication between the two models. Both the DSGE prior and the BVAR prior play an integral part of model estimation. Moreover, the two likelihoods interact with each other in forming the merged likelihood.

Cogley and Sargent (2005) study an economy in which agents, facing model uncertainty, compute the posterior odds ratios over three models and make decisions by Bayesian model averaging. As pointed out by Sims (2003) and Geweke and Amisano (forthcoming), one can encounter the pathology that the odds ratios lead to selecting only one model and rejecting all other models. By estimating the state-dependent weights and the parameters of the models jointly, we provide an empirically operational

[3]West and Harrison (1997) present a similar idea of allowing the weights of each model time-varying in dynamic forecasting exercises.

way to implement Sims (2003)'s idea of "filling in the gap" between DSGE and BVAR models by overcoming the difficulties inherent in Bayesian model averaging.[4]

III. Markov-switching framework

To integrate model uncertainty and parameter uncertainty in one merged framework, we propose a Bayesian approach to modeling state-dependent weights for a linear combination of predictive densities produced by different models. Our key assumption is that the observed data at time t, y_t^o, is generated from the following predictive density

$$p\left(y_t \mid Y_{t-1}^o, \Theta, Q, w\right) = \sum_{i=1}^{n} w_{i,t}^* p\left(y_t \mid Y_{t-1}^o, \Theta_i, M_i\right),$$

where

$$w_{i,t}^* = \sum_{s_t=1}^{h} w_{i,s_t} p\left(s_t \mid Y_{t-1}^o, \Theta, Q, w\right),$$

$p\left(y_t \mid Y_{t-1}^o, \Theta_i, M_i\right)$ is the predictive density of y_t conditional on the parameters of model i and the observed data up to time $t-1$, $Y_{t-1}^o = \left\{y_1^o, \cdots, y_{t-1}^o\right\}$, Θ_i is a set of parameters for model i, and w_{i,s_t} is the probability weight given to model i when the state, s_t, occurs at time t with $w_{i,s_t} \geq 0$ and $\sum_{i=1}^n w_{i,s_t} = 1$. The state variable, s_t, follows a Markov process with the transition matrix Q, where $\text{Prob}\left[s_t = k \mid s_{t-1} = j\right] = q_{k,j}$ for $k, j = 1, \ldots, h$.[5] Note that

$$\Theta = \{\Theta_1, \cdots, \Theta_n\},\ w = \{w_{i,k}\} \text{ for } k = 1, \ldots, h, i = 1, \ldots, n.$$

We use the notation "M_i" in $p\left(y_t \mid Y_{t-1}^o, \Theta_i, M_i\right)$ because we will compare the marginal data density of model i, denoted by $p\left(Y_T^o \mid M_i\right)$, with the marginal data density of the complete model, denoted by $p\left(Y_T^o \mid M\right)$.

The log likelihood function is thus given by

$$\log p\left(Y_T^o | \Theta, Q, w\right) = \sum_{t=1}^{T} \log p\left(y_t \mid Y_{t-1}^o, \Theta, Q, w\right) =$$
$$\sum_{t=1}^{T} \log\left[\sum_{i=1}^{n}\left(\sum_{s_t=1}^{h} w_{i,s_t} p\left(s_t \mid Y_{t-1}^o, \Theta, Q, w\right)\right) p\left(y_t^o \mid Y_{t-1}^o, \Theta_i, M_i\right)\right],$$

where the parameters Θ, Q, and w are to be *estimated jointly*.

[4] Hansen and Sargent (2001) and Sims (2003) advocate a large model space. In our framework, this advice corresponds to increasing the number of individual models in the merged model space.

[5] As shown in Sims, Waggoner, and Zha (2008), q_{kj} can also depend on the observed data Y_{t-1}^o.

Whether model i will be preferred by the data depends on the state s_t. We use the random variable $\xi_t \in \{1, \ldots, n\}$ to index the model chosen at time t. This random variable obeys the conditional probability:

$$p(\xi_t \mid s_t) = w_{s_t, \xi_t},$$

where $\sum_{\xi_t=1}^{n} p(\xi_t \mid s_t) = 1$ because $\sum_{\xi_t=1}^{n} w_{s_t,\xi_t} = 1$. Although ξ_t is not a Markov-process itself, the joint process (s_t, ξ_t) is.

Proposition 1. The joint process (s_t, ξ_t) is a Markov process with the expanded transition matrix

$$\text{Prob}\left[(s_t, \xi_t) = (k, i) \mid (s_{t-1}, \xi_{t-1}) = (j, g)\right] = q_{k,j} w_{i,k},$$

for $j, k = 1, \ldots, h$ and $g, i = 1, \ldots, n$.

Proof. The proof follows from the basic conditional probability theory by noting that

$$p(s_t, \xi_t \mid s_{t-1}, \xi_{t-1}) = p(s_t \mid s_{t-1}, \xi_{t-1})\, p(\xi_t \mid s_t, s_{t-1}, \xi_{t-1}) = p(s_t \mid s_{t-1})\, p(\xi_t \mid s_t).$$

□

Proposition 1 formulates the way we implement our estimation strategy. Since the restrictions imposed on the expanded transition matrix in Proposition 1 satisfy the conditions specified in Sims, Waggoner, and Zha (2008), one can apply their estimation method directly to our framework of merging individual models.

IV. Identification and reinterpretation

In this section we discuss the identification of state-dependent weights and reinterpret what constant weights used in the literature mean from the ex ante point of view.

IV.1. **General identification issue.** In general, $w_{i,k}$ and $p\left(s_t \mid Y_{t-1}^o, \Theta, Q, w\right)$ can be identified separately. To see this point, consider the following case with $h = 2$ and $n = 2$:

$$\begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} \begin{bmatrix} p\left(s_t = 1 \mid Y_{t-1}^o, \Theta, Q, w\right) \\ p\left(s_t = 2 \mid Y_{t-1}^o, \Theta, Q, w\right) \end{bmatrix} = \begin{bmatrix} w_{1,t}^* \\ w_{2,t}^* \\ w_{3,t}^* \end{bmatrix}.$$

For any given $w_{i,t}^*$ and $p\left(s_t \mid Y_{t-1}^o, \Theta, Q, w\right)$, there are three equations but four unrestricted weights $w_{i,k}$ and it appears that we always have more than one solution. This conclusion is not true, however. Since both $w_{i,t}^*$ and $p\left(s_t \mid Y_{t-1}^o, \Theta, Q, w\right)$ change over time but $w_{i,k}$ are constant, we do not have more than one solution and may indeed have no solution at all for some t or s_t. This results means that we cannot arbitrarily change $w^*_{i,t}$ and $p\left(s_t \mid Y^o_{t-1}, \Theta, Q, w\right)$ while keeping $w_{i,k}$ the same across time. Thus, we can in general identify w_{i,s_t}.

IV.2. **Strengthening identification.** As the number of models or the number of states increases, the number of free parameters in the expanded transition matrix increases at an even faster speed, making it necessary to impose further restrictions to avoid overfitting and at the same time strengthen the identification of w_{i,s_t}. To achieve this goal, we let $h = n$, $w_{i,s_t} = 1$ when $s_t = i$, and $w_{j,s_t} = 0$ when $s_t \neq j$. Thus, when the state s_t is realized, only one of the models is operative. Since one can never be sure of which state is realized, one can never be sure of which model is operative, even after observing all the data. One can, however, compute the smoothed probability of the state, $p\left(s_t \mid Y^o_T, \hat{\Theta}, \hat{Q}, \hat{w}\right)$, where the superscript $\hat{}$ denotes the posterior estimate. The probability enables one to gauge how likely a particular model is selected. In our application, we will report this posterior probability throughout the history.

IV.3. **Reinterpretation.** Although we know which model is operative given the current state s_t, there is uncertainty about models ex ante (i.e., at time $t-1$) and forecasts of economic variables will in general depend on multiple models through the transition matrix. Thus, for the purpose of policy forecasts, it is *ex ante* uncertainty that matters. Moreover, this uncertainty presents a different interpretation of constant weights used in the literature, as shown in the following proposition.

Proposition 2. If $q_{i,j} = q_{i,k} = q_i$ for $i, j, k = 1, \ldots, n$, it must be true that $w^*_{i,t} = q_i$.

Proof. Because $q_{i,j} = q_{i,k}$, the probability of switching to the current state s_t is the same no matter what the state at time $t-1$ is. This result means that all the past data are irrelevant in inferring about the probability of the current state. It follows that

$$p\left(s_t = i \mid Y^o_{t-1}, \Theta, Q, w\right) = q_i.$$

From the definition of $w^*_{i,t}$, we have

$$w^*_{i,t} = \sum_{s_t=1}^{h} w_{i,s_t} p\left(s_t \mid Y^o_{t-1}, \Theta, Q, w\right) = w_{i,i} p\left(s_t = i \mid Y^o_{t-1}, \Theta, Q, w\right) = q_i.$$

□

It is, perhaps, not surprising that constant weights are a special case of our Markov-switching framework. What is new from Proposition 2 is that a constant weight is about the relative importance of the model only at time $t-1$ and the model's weight will change once we have the data beyond time $t-1$. Given all the data, moreover, our Markov-switching framework enables us to reinterpret this history by tracing out the periods in which a particular model is more relevant than others, even when all the weights are constant.

V. Application

We apply the framework presented in Section III to two widely used models: a medium-scale DSGE model and a BVAR model. The DSGE model is based on Liu, Waggoner, and Zha (2010). The large part of the model is the same as Altig, Christiano, Eichenbaum, and Linde (2004) and Smets and Wouters (2007) with the notable exceptions that (1) some real rigidity is introduced, as in Chari, Kehoe, and McGrattan (2000), by assuming the existence of firm-specific factors (such as land) such that the sum of cost shares of capital and labor inputs is less or equal to one and (2) a shock to the depreciation in physical capital is introduced as a stand-in for economic obsolescence of capital (see Appendix B for some details of the model).

The DSGE model is fit to eight quarterly variables: quarterly growth of real per capita GDP ($\Delta \log Y_t^{\mathrm{Data}}$), quarterly growth of real per capita consumption ($\Delta \log C_t^{\mathrm{Data}}$), quarterly growth of real per capita investment in capital goods unit ($\Delta \log I_t^{\mathrm{Data}}$), quarterly growth of the real wage ($\Delta \log w_t^{\mathrm{Data}}$), the quarterly GDP-deflator inflation rate (π_t^{Data}), quarterly growth of per capita hours ($\Delta \log L_t^{\mathrm{Data}}$), the federal funds rate ($\mathrm{FFR}_t^{\mathrm{Data}}$), and quarterly growth of investment-specific technology ($\Delta \log Q_t^{\mathrm{Data}}$) as measured by the inverse of the relative price of investment. A detailed description of the data is given in Appendix A. The data in the initial four quarters from 1960:I to 1960:IV are used to obtain the initial condition at 1961:I for the Kalman filter. Thus, the effective sample used for model evaluation is from 1961:I to 2010:II.

The BVAR model has the same eight variables as the DSGE model; and it has four lags from 1960:I to 1960:IV so that the effective sample is also from 1961:I to 2010:II. To make our BVAR model comparable with the DSGE literature, we follow Smets and Wouters (2007) and use the standard "Minnesota-like" prior with the hyperparameter values $\mu_1 = \mu_2 = \mu_3 = 1.5$, and $\mu_4 = 1.3$ where μ_1 controls overall tightness of the random walk prior, μ_2 controls relative tightness of the random walk prior on the lagged coefficients, μ_3 controls relative tightness of the random walk prior on the constant term, and μ_4 controls tightness of the prior that dampens the erratic sampling effects on lag coefficients (lag decay).[6]

[6]In Section VI.4, we study another standard prior proposed by Sims and Zha (1998).

The prior for the DSGE model is reported in Tables 1 and 2. Instead of specifying the mean and the standard deviation, we use the 90% probability interval to back out the hyperparameter values of the prior distribution. The intervals are generally set wide enough to allow for the possibility that the posterior mode is close to or on the boundary of the parameter space. It also allows for multiple local posterior peaks (Del Negro and Schorfheide, 2008). Our approach is necessary to deal with skewed distributions and allows for reasonable hyperparameter values in certain distributions, such as the Inverse-Gamma, where the first two moments may not exist.

For many parameters with the Beta prior distribution, such as the habit parameter and the persistence parameters in shock processes, we insist on a positive probability density at the value 0 to allow for the possibility of no habit and no persistence at all; we also insist on zero probability density at the value 1 to maintain the assumption that the economy is on the balanced growth path. Consequently, the two hyperparameter values for the Beta prior are set at 1.0 and 2.0.

The prior for the labor share and capital share is the Beta distribution with the restriction $\alpha_1 + \alpha_2 \leq 1$ such that the production technology requires firm-specific factors (Chari, Kehoe, and McGrattan, 2000). The bounds for the α_1 values in the 90% probability interval are 0.3 and 0.4 and those for α_2 are 0.5 and 0.7. With the restriction $\alpha_1 + \alpha_2 \leq 1$, however, the joint 90% probability region would be somewhat different.

The prior for the inverse Frisch elasticity η follows the Gamma distribution. We choose the 2 hyper-parameters of the Gamma distribution such that the lower bound (0.2) and the upper bound (10.0) of η correspond to the 90% probability interval. This prior range for η implies that the Frisch elasticity lies between 0.1 and 5.

The lower and upper bounds of prior distributions are specified in Table 1 for the parameters λ_q, λ_*, β, σ_u, S'', δ, ξ_p, γ_p, ξ_w, γ_w, ϕ_π, ϕ_y, and π^*. Using these wide bounds, we back out the two hyperparameter values for the corresponding prior distributions.

The Gamma prior for the average net price markup $\mu_p - 1$ is the same as the Gamma prior for the average net wage markup $\mu_w - 1$. By setting the first hyperparameter of this prior to be 1.0, we allow for a positive probability that the net markups may be zero. This generality (a less stringent prior) turns out to be critical as our posterior estimates of $\mu_p - 1$ and $\mu_w - 1$ are nearly zero. We set the second hyperparameter of the Gamma prior at 5.5 such that the implied 90% probability bounds are wide enough (from 0.0094 to 0.5446).

The prior for the parameter ρ_{gz}, capturing the impact of technological improvement on government spending, is the Gamma distribution with the 90% probability bounds given by $[0.2, 3.0]$.

The standard deviation of each of the 8 shocks has the Inverse Gamma prior distribution with the 90% probability bounds given by $[0.0005, 1.0]$. These wide bounds are necessary to take account of the possibility that some shocks may have very small variances while others may have very large variances. With these bounds, there exist no moments for the Inverse Gamma prior. One can still, however, back out the two hyperparameter values as reported in Table 2.

The transition from one model to the other has the following matrix form:

$$Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix},$$

where $\sum_{i=1}^{2} q_{ij} = 1$ for $j = 1, 2$. Following Sims, Waggoner, and Zha (2008), we express a prior belief that the average duration for a model to remain dominant is between six and seven quarters. The belief implies that the hyperparameter in the exponent of q_{ii} in the Dirichlet prior density is 5.6667 and the other hyperparameter is 1.0. This prior setting allows for the possibility that model i dominates other models all the time (i.e., $q_{ii} = 1$). Table 2 reports the corresponding 90% probability interval.

VI. Measuring misspecification

In this section we quantify the degree of DSGE model misspecification by 1) computing the MDDs for the DSGE and BVAR models against the MDD for the merged model and 2) tracing out the posterior probabilities of each model across time. We then discuss a variety of economic implications of this misspecification. Although both BVAR and DSGE models are misspecified, we focus on the DSGE model by comparing the estimated results of the merged model to those of the DSGE model alone.

VI.1. **Model fit.** We compute the MDDs for the merged model, the DSGE model alone, and the BVAR model alone. Table 5 reports log values of these MDDs. For the BVAR model, there is an analytical solution for calculating the MDD so that the reported log value of MDD has negligible numerical errors. For the DSGE model and the merged model, however, numerical errors are nontrivial. We use two different Monte Carlo methods to compute MDDs. One method is the truncated modified harmonic mean (MHM) method proposed by Sims, Waggoner, and Zha (2008); the other method, called the "Müeller method," is developed by Ulrich Müeller at Princeton University.[7] The two methods can give different results due to numerical errors and we report the range of estimates of the MDDs in Table 5.[8]

The log value of the MDD for the DSGE model is about 50 above log MDD for the BVAR. The conventional Bayesian averaging procedure would give the BVAR essentially zero weight. The merged model, unlike the conventional Bayesian averaging procedure, not just combines the two distinct models but also expands the parameter space by estimating the parameters of both models and the weights jointly. Consequently, both models are operative as discussed in Section VI.2. The resulting MDD for the merged model is about 100 in log value above the MDD for the DSGE model. This magnitude gives a sense of how misspecified both models are.

VI.2. **Posterior estimates.** The prior specified for the DSGE model is looser and more agnostic than most priors in the DSGE literature. The agnostic prior comes also with the price: since the likelihood function for the merged model is complicated and full of multiple local peaks, the resulting posterior density function is complicated as well. The non-Gaussian nature of the posterior density implies that the posterior mean may have a very low (joint) probability and thus cannot represent the most likely outcome for the model. The posterior mode is, by definition, the most probable point in the parameter space, regardless of how non-Gaussian and complicated the shape of the posterior probability density is. Moreover, using a point in the neighborhood of the posterior mode as a starting point for the MCMC algorithm avoids the situation where a long sequence of posterior draws gets stuck in the low probability region due to a poor starting point.

To find the posterior mode, we combine the hill-climbing quasi-Newton (Broyden-Fletcher-Goldfarb-Shanno – BFGS) algorithm with occasional downhill movements generated by MCMC draws. Tables 3 and 4 report the posterior-mode estimates of the DSGE model parameters along with the 90% marginal probability intervals. In these tables we contrast the estimated results for the merged model to those for the DSGE model alone. There are a few instances in which the estimated results from the merged model are similar to those from the DSGE model when estimated alone. The probability interval of β is actually smaller in the merged model than in the DSGE

[7]See Liu, Waggoner, and Zha (2010) for a detailed description of the Müeller method.

[8]To ensure the accuracy, 20 million posterior draws and 2 million proposal draws are simulated. For the merged model, the simulation takes about 30 days or two full days by availing itself to computational parallelism on a cluster of 15 modern computers.

model alone. The estimate of the average price markup is close to zero, similar to the estimate in the DSGE model when treated alone. This result implies that the demand curve for differentiated goods is very flat. Thus, a small increase in the relative price can lead to large declines in relative output demand. Even if firms can re-optimize their pricing decisions frequently, they choose not to adjust their relative prices too much. In other words, the small average markup and thus the large demand elasticity become a source of strategic complementarity in firms' pricing decisions.

The general pattern, as indicated by the 90% probability intervals, is that the merged model exposes more uncertainty about the estimated DSGE parameters than what is implied when the DSGE model is treated as the truth and estimated alone. In many cases, such as the inverse Frisch elasticity of labor supply (η) and the curvature of the capital utilization cost function evaluated at the steady state (σ_u), the probability distributions have changed so much that the posterior estimates are very different. The inflation target (π^*) is another example in point. Our prior on this parameter is very loose, covering the range from 1% to 8% for the annualized rates (Table 1). The marginal posterior distribution for π^* is very wide for both the DSGE model and for the merged model, but the distribution for the merged shifts to the left and gives a substantial probability (more than 45%) to the target below 4%.[9] The estimate of the capital share α_1 has increased and the estimate of the labor share α_2 has decreased so that the sum $\alpha_1 + \alpha_2$ in the merged model is considerably smaller than that in the DSGE model, implying that this source of real rigidity is strong.

Perhaps most notable changes pertain to some persistence parameters. As shown in Table 4, the 90% probability intervals for the parameters ρ_p, ϕ_p, and ρ_a are much wider in the merged model than in the DSGE model alone. The posterior distributions for persistence parameters tend to have a long fat tail toward zero, indicating much more uncertainty about the highly persistent shock processes than the DSGE model would recommend.

Remember that a combined number of parameters from the two models is very large and the shape of the posterior probability density over this high-dimensional parameter space is extremely non-Gaussiann full of skewness and fat-tails. When we compute the marginal 90% probability interval of one parameter by *integrating out all the rest of the parameters*, it is not uncommon that some posterior mode estimates fall outside the 90% probability intervals as indicated in Tables 3 and 4. Take the two parameters

[9]Our sample covers the several high inflation periods. The estimated target is much lower if we use only the sample after 1987.

η and ϕ_w as an example. The posterior-mode estimates of these two parameters are outside the corresponding *marginal* 90% probability intervals. Figure 1 plots the two-dimensional joint probability density function of η and ϕ_w. It can be seen from the figure that the shape of this distribution has a mass probability density around the boundary defined by $\eta = 0$ and $\phi_w = 0$ coupled with fat long tails. Since this two-dimensional probability density has already been marginalized by integrating out the other hundreds of parameters in the merged model, it gives us only a glimpse of the complexity of the shape of the high-dimensional joint probability density, which is beyond visualization.

The resultant disagreement between the joint distribution and a marginal distribution also shows up in the estimate and inference of q_{11}, which measures the duration in which the DSGE model dominates the BVAR model. The posterior-mode estimate of $q_{1,1}$ is outside the 90% probability interval and the marginal distribution of q_{11} is clearly skewed to the right. The estimate of $q_{1,1}$ is 0.309, implying that the duration in which the DSGE model dominates the BVAR model is about 1.5 quarters. As judged by the 90% probability interval, the duration is unlikely to last for more than 3 quarters. On the other hand, the estimate of $q_{2,2}$ is 0.72 and thus the most likely duration in which the BVAR model dominates the DSGE model is about 3.5 quarters. The duration can last as long as 7 quarters, as determined by the upper bound of the 90% interval (Table 4).

VI.3. **A historical perspective of the role of a model.** The transition probability, $q_{i,i}$, measures the average (unconditional) importance of model i. Often one is interested in knowing how important model i is at a particular time of the history. Figure 2 displays the posterior probabilities of the DSGE model. Clearly, the DSGE model is operative throughout the history, but for the most part, the probability of the DSGE model being near one lasts no more than one quarter at a time, consistent with the estimate reported in Table 4. Moreover, the estimated DSGE model performs poorly during the recessions, as indicated by the shaded bars in Figure 2.

In contrast, the probability of the BVAR model near one (i.e., the probability of the DSGE model near zero in Figure 2) tends to last for a few quarters at a time.

The result that the DSGE model is operative sporadically throughout the history can be partially explained by Figure 3, which displays the log values of predictive densities of the merged model, the DSGE model, and the BVAR model. Clearly the merged model has higher predictive densities than both the DSGE and BVAR models throughout the entire history. The times when the predictive density of the DSGE model is higher than the BVAR model are irregular and scattered without much duration. Although the MDD for the DSGE model is much higher than the MDD for the BVAR model, the data prefers the DSGE model only intermittently throughout the sample.

VI.4. **Prior sensitivity.** The MDD of a particular model is very sensitive to prior specifications. In particular, the BVAR model has hundreds of parameters and the MDD varies wildly with different priors. The "Minnesota-like" prior used in Smets and Wouters (2007) ignores cross effects among variables and the correlation between the constant term and other coefficients. Sims and Zha (1998) introduces additional dummy-observation components of the prior that incorporate correlations in prior beliefs about all coefficients (including the constant term) in every equation. Thus, the model is pulled toward a form in which either all variables are stationary with means equal to the sample averages of the initial conditions or there are cointegration relationships.

The Sims and Zha (1998) prior has been found to improve out-of-sample forecasts in a variety of contexts with economic time series. Indeed, when we use the exact prior recommended by Sims and Zha (1998), the log MDD of the BVAR is increased to 5894.6, as compared to 5685.7 in Table 5. This MDD is about 150 in log value higher than the DSGE counterpart (Table 5). Given this stark fact, one might conclude that the DSGE model must play no or little role in the merged model space. This conclusion would be incorrect. The resultant merged model has the log value of MDD being in the range from 6039.0 to 6044.4. The MDD of the merged model is much higher than the MDD of the BVAR, because the DSGE model continues to form an integral part of the model space in fitting the data. The posterior estimate of $q_{1,1}$ rises to 0.473, while the posterior estimate of $q_{2,2}$ rises to 0.833. Moreover, the posterior probabilities of the DSGE model throughout the history have a pattern similar to Figure 2.

In general, when the prior specification for an individual model changes, the MDD can change drastically. But our extensive experiments indicate that the merged model pooling together the two models is insensitive to changes in prior specifications, in the sense that it dominates individual models by allowing both models to form an integral part of the data generating process.

VII. Economic implications

We are now in a position to discuss economic implications when one takes explicit account of both model uncertainty and parameter uncertainty in our merged framework.

VII.1. **Output fluctuations.** A shock to capital or investment, such as a capital depreciation shock, plays an important role in output fluctuations. Table 6 shows that contributions from the capital depreciation shock account for close to 50% of fluctuations in output in the short run (within two years) and about 40% of output fluctuations in the longer run (for three to five years). The DSGE model, if it is treated in isolation, would underestimate the magnitude of the contributions from the capital depreciation shock in output fluctuations. The underestimation is at least by 10 percentage points for most forecast horizons, as reported in Table 6.

VII.2. **Posterior distributions.** Figure 4 displays the marginal posterior distributions of four key structural parameters from the merged model (left hand column) and the DSGE model alone (right hand column). The posterior distributions from the merged model uncover considerably more uncertainty about the parameters than what is implied by the DSGE model alone. Moreover, the posterior distributions shift, giving more probability to the untrodden regions.

- For the inflation coefficient in the Taylor rule (ϕ_π), the merged model puts almost zero probability on the value below 1.5 , while the DSGE model in isolation would put mass probability around 1.5 with a considerably tighter 90% probability interval.
- For the Calvo price parameter (ξ_p), the posterior distribution from the merged model shifts to the right, giving substantial probability to the values between 0.6 and 0.8 as well as between 0.1 and 0.4.
- For the Calvo wage parameter (ξ_w), the posterior distribution from the merged model shifts to the left, giving considerable probability to the values between between 0.1 and 0.6, whereas the posterior distribution from the DSGE model estimated in isolation concentrates around 0.4 with a much tighter 90% probability interval.
- For the parameter (S'') measuring investment adjustment costs, the posterior distribution from the merged model spreads out to the values beyond 2, indicating that the higher investment adjustment costs (between 2 and 4) is probable.

Our estimates show that the estimation of the DSGE model utilizes roughly one third of the data points in the sample. It is unsurprising that the error bands of DSGE parameters are wider for the merged model. What is new in our findings, however, is that the error bands in the merged model are much more than 1.73 (a square root of three) times those when the DSGE model is estimated alone with all the data points. Figure 2 provides an insight of our findings. Since the DSGE model dominates the BVAR model only for the periods in which the data have more similarity than the data in other periods, the data that experience large fluctuations (as in the recession periods) are excluded in the estimation of DSGE parameters. This exclusion results in considerably more uncertainty about the estimates than what the number of data points would suggest.

VII.3. **Dynamic responses.** Figure 5 shows the impulse responses of output, consumption, real wage, and inflation to a one-standard-deviation shock to capital depreciation. The left hand column shows the responses generated from the estimated merged model and the right hand column shows the responses from the DSGE model when it is estimated in isolation. Comparing the two columns side by side, one can see the notable differences between the merged model and the DSGE model.

- Output responses in the merged model are very persistent, while the corresponding responses in the DSGE model alone return to the steady state after two and a half years.
- The magnitude of consumption and real wage responses in the merged model is considerably larger than that in the DSGE model when it is estimated separately.
- A shock to capital depreciation is a negative shock to the capital stock and thus the agent's wealth. As a result, consumption falls due to the wealth effect, but the marginal cost of capital rises due to the decline in the capital stock. When the DSGE model is estimated in isolation, the rise in the marginal cost of capital slightly dominates the fall in the real wage. Thus, the increase in inflation responses is significant statistically but the magnitude is insignificant economically. In the merged model, however, the fall in the real wage overweighs the rise in the marginal cost of capital so that inflation fall. In contrast to the results generated from the DSGE model alone, inflation responses are predominantly negative in the short run (within the two years) before they rise in the longer run (after the third year).

Similar to the findings discussed in previous sections, the error bands of impulse responses in the merged model (left hand column in Figure 5) are considerably wider than those generated by giving the DSGE model all the weight. These results emphasize the underlying uncertainty ignored by discarding the BVAR model in the model space.

VIII. Conclusion

When a particular model is usable for policy prescriptions, economists understand that the model is an approximation at best and should be used only with a grain of salt. A positive question is how to quantify the degree to which the model is misspecified. Using a structural DSGE model and a reduced-form BVAR model as an economic laboratory, we demonstrate that a merger of the two models exposes how misspecified both models are. In particular, we show that even though the MDD for the DSGE model is much higher than the MDD for the BVAR model, the DSGE model dominates the BVAR model sporadically for only one third of the history. The estimated results from the merged model significantly alter the economic implications derived from the DSGE parameters and their impulse responses.

The framework studied in this paper is general enough to be applicable to a variety of economic questions beyond the particular application used in this paper. One can, for example, study a structural BVAR model by identifying economic shocks such as a monetary policy shock, a credit shock, an oil price shock, and a technology shock. One can then merge this structural BVAR model with the DSGE model that has the same set of economic shocks. The formal communication between these two structural models, facilitated by our framework, allows the researcher to reconcile the differences between impulse responses implied by two isolated models when they are estimated separately. Moreover, the approach explored in this paper allows for more than two models, and the models included in the merged framework need not be nested.

Appendix A. Detailed data description

All data are constructed from the original data in the Haver Analytics Database. The constructed data, the original data identifiers, and the data sources are described below.

- $Y_t^{\text{Data}} = \frac{\text{GDPH}}{\text{LN16N@USECON}}$.
- $C_t^{\text{Data}} = \frac{\text{(CN@USECON + CS@USECON - CSRU@USECON)*100/JCXFE@USNA}}{\text{LN16N@USECON}}$.
- $I_t^{\text{Data}} = \frac{\text{(CD@USECON + FNE@USECON)*100/JCXFE@USNA}}{\text{LN16N@USECON}}$.

- $w_t^{\text{Data}} = \frac{\text{LXNFC@USECON/100}}{\text{JCXFE@USNA}}$.
- $\pi_t^{\text{Data}} = \frac{\text{JCXFE@USNA}_t}{\text{JCXFE@USNA}_{t-1}}$.
- $L_t^{\text{Data}} = \frac{\text{LXNFH@USECON}}{\text{LN16N@USECON}}$.
- $\text{FFR}_t^{\text{Data}} = \frac{\text{FFED@USECON}}{400}$.
- $Q_t^{\text{Data}} = \frac{\text{JCXFE@USNA}}{\text{GordonPriceCDplusES}}$.

LN16N@USECON: Civilian noninstitutional population: 16 years and over. Breaks in population are eliminated from 10-year censuses and post 2000 American Community Surveys using "error of closure" method. This fairly simple method was used by the Census Bureau to get a smooth population monthly population series. This smooth series reduces the unusual influence of drastic demographic changes. Source: BLS.

GDPH: Real gross domestic product (2005 dollars). Source: BEA.

CN@USECON: Nominal personal consumption expenditures: nondurable goods. Source: BEA.

CS@USECON: Nominal consumption expenditures: services. Source: BEA.

CSRU@USECON: Nominal personal consumption expenditures: housing and utilities. Source: BEA.

CD@USECON: Nominal personal consumption expenditures: durable goods. Source: BEA.

FNE@USECON: Nominal private nonresidential investment: equipment & software. Source: BEA.

JCXFE@USNA: PCE excluding Food and Energy: Chain Price Index (2005=100). Source: BEA.

LXNFC@USECON: Nonfarm business sector: compensation per hour (1992=100). Source: BLS.

LXNFH@USECON: Nonfarm business sector: hours of all persons (1992=100). Source: BLS.

FFED@USECON: Nnnualized federal funds effective rate. Source: FRB.

GordonPriceCDplusES: Investment deflator. The Tornquist procedure is used to construct this deflator as a weighted aggregate index from the four quality-adjusted price indexes: private nonresidential structures investment, private residential investment, private nonresidential equipment & software investment, and personal consumption expenditures on durable goods. Each price index is a weighted one from a number of individual price series within this categories. For each individual price series from 1947 to 1983, we use Gordon (1990)'s quality-adjusted price index. Following Cummins and Violante (2002), we estimate an econometric model of Gordon's price series as a function of a time trend and a few NIPA indicators (including the current and lagged values of the corresponding NIPA price series); the estimated coefficients are then used to extrapolate the quality-adjusted price index for each individual price series for the sample from 1984 to 2007. These constructed price series are annual. Denton (1971)'s method is used to interpolate these annual series on a quarterly frequency. The Tornquist procedure is then used to construct each quality-adjusted price index from the appropriate interpolated quarterly price series.

Appendix B. DSGE equilibrium dynamics

We introduce the notation $\Delta x_t = x_t - x_{t-1}$. We use the hat variable, $\hat{x}_t$, to denote the log deviation of the stationary variable X_t from its steady state value (i.e., $\hat{x}_t = \log(X_t/X)$). The log-linearized equilibrium conditions for our DSGE mode, below, summarize the equilibrium dynamics.

$$
\begin{aligned}
\hat{\pi}_t - \gamma_p \hat{\pi}_{t-1} &= \frac{\kappa_p}{1+\bar{\alpha}\theta_p}(\hat{\mu}_{pt} + \hat{mc}_t) + \beta \mathrm{E}_t[\hat{\pi}_{t+1} - \gamma_p \hat{\pi}_t], \text{ (price-Phillips curve)} && \text{(A1)} \\
\hat{w}_t - \hat{w}_{t-1} &+ \hat{\pi}_t - \gamma_w \hat{\pi}_{t-1} = \frac{\kappa_w}{1+\eta\theta_w}(\hat{\mu}_{wt} + m\hat{r}s_t - \hat{w}_t) + \\
& \beta \mathrm{E}_t[\hat{w}_{t+1} - \hat{w}_t + \hat{\pi}_{t+1} - \gamma_w \hat{\pi}_t], \text{ (wage-Phillips curve)} && \text{(A2)} \\
\hat{q}_{kt} &= S'' \lambda_I^2 \left\{ \Delta \hat{i}_t + \frac{1}{1-\alpha_1}(\Delta \hat{q}_t + \alpha_2 \Delta \hat{z}_t) \right. \\
& \left. - \beta \mathrm{E}_t \left[\Delta \hat{i}_{t+1} + \frac{1}{1-\alpha_1}(\Delta \hat{q}_{t+1} + \alpha_2 \Delta \hat{z}_{t+1}) \right] \right\}, \text{ (investment decision)} && \text{(A3)} \\
\hat{q}_{kt} &= \mathrm{E}_t \left\{ \Delta \hat{a}_{t+1} + \Delta \hat{U}_{c,t+1} - \frac{1}{1-\alpha_1}\left[\alpha_2 \Delta \hat{z}_{t+1} + \Delta \hat{q}_{t+1}\right] \right. \\
& \left. + \frac{\beta}{\lambda_I}\left[(1-\delta)\hat{q}_{k,t+1} - \delta \hat{\delta}_{t+1} + \tilde{r}_k \hat{r}_{k,t+1}\right] \right\}, \text{ (capital decision)} && \text{(A4)} \\
\hat{r}_{kt} &= \sigma_u \hat{u}_t, \text{ (capacity utilization)} && \text{(A5)} \\
0 &= \mathrm{E}_t \left[\Delta \hat{a}_{t+1} + \Delta \hat{U}_{c,t+1} \right. \\
& \left. - \frac{1}{1-\alpha_1}\left[\alpha_2 \Delta \hat{z}_{t+1} + \alpha_1 \Delta \hat{q}_{t+1}\right] + \hat{R}_t - \hat{\pi}_{t+1} \right], \text{ (bond decision)} && \text{(A6)} \\
\hat{k}_t &= \frac{1-\delta}{\lambda_I}\left[\hat{k}_{t-1} - \frac{1}{1-\alpha_1}(\alpha_2 \Delta \hat{z}_t + \Delta \hat{q}_t) \right] \\
& - \frac{\delta}{\lambda_I}\hat{\delta}_t + \left(1 - \frac{1-\delta}{\lambda_I}\right)\hat{i}_t, \text{ (capital law of motion)} && \text{(A7)} \\
\hat{y}_t &= c_y \hat{c}_t + i_y \hat{i}_t + u_y \hat{u}_t + g_y \hat{g}_t, \text{ (resource constraint)} && \text{(A8)}
\end{aligned}
$$

$$\hat{y}_t = \alpha_1 \left[\hat{k}_{t-1} + \hat{u}_t - \frac{1}{1-\alpha_1}(\alpha_2 \Delta \hat{z}_t + \Delta \hat{q}_t) \right] + \alpha_2 \hat{l}_t, \text{ (production function)} \tag{A9}$$

$$\hat{w}_t = \hat{r}_{kt} + \hat{k}_{t-1} + \hat{u}_t - \frac{1}{1-\alpha_1}(\alpha_2 \Delta \hat{z}_t + \Delta \hat{q}_t) - \hat{l}_t, \text{ (labor \& capital demand)} \tag{A10}$$

$$\hat{R}_t = \rho_r \hat{R}_{t-1} + (1-\rho_r)\left[\phi_\pi \hat{\pi}_t + \phi_y \hat{y}_t\right] + \sigma_r \varepsilon_{rt}, \text{ (interest rate rule)} \tag{A11}$$

where

$$\hat{mc}_t = \frac{1}{\alpha_1 + \alpha_2}[\alpha_1 \hat{r}_{kt} + \alpha_2 \hat{w}_t] + \bar{\alpha}\hat{y}_t, \tag{A12}$$

$$\hat{mrs}_t = \eta \hat{l}_t - \hat{U}_{ct}, \tag{A13}$$

$$\begin{aligned} \hat{U}_{ct} = {} & \frac{\beta b(1-\rho_a)}{\lambda_* - \beta b}\hat{a}_t - \frac{\lambda_*}{(\lambda_* - b)(\lambda_* - \beta b)}[\lambda_* \hat{c}_t - b(\hat{c}_{t-1} - \Delta \hat{\lambda}_t^*)] \\ & + \frac{\beta b}{(\lambda_* - b)(\lambda_* - \beta b)}[\lambda_* \mathrm{E}_t(\hat{c}_{t+1} + \Delta \hat{\lambda}_{t+1}^*) - b\hat{c}_t], \end{aligned} \tag{A14}$$

Note that $\hat{\pi}_t$ is inflation, $\hat{w}_t$ is real wage, $\hat{q}_{kt}$ is the shadow price of existing capital (Tobin's q), $\hat{i}_t$ is investment, $\hat{q}_t$ is the biased technology shock process, $\hat{z}_t$ is the neutral technology shock process, $\hat{a}_t$ is the risk premium (preference) shock process, $\hat{u}_t$ is the utilization rate of capital, $\hat{r}_{kt}$ is the real rental price of capital, $\hat{\delta}_t$ is the capital depreciation shock process, $\hat{R}_t$ is the nominal rate of interest, $\hat{k}_t$ is the capital stock, $\hat{y}_t$ is output, $\hat{c}_t$ is consumption, $\hat{g}_t$ is government spending, and $\hat{l}_t$ is hours worked.

The steady-state variables are given by

$$\tilde{r}_k = \frac{\lambda_I}{\beta} - (1-\delta), \tag{A15}$$

$$u_y \equiv \frac{\tilde{r}_k \tilde{K}}{\tilde{Y}\lambda_I} = \frac{\alpha_1}{\mu_p}, \tag{A16}$$

$$i_y = [\lambda_I - (1-\delta)]\frac{\alpha_1}{\mu_p \tilde{r}_k}, \tag{A17}$$

$$c_y = 1 - i_y - g_y. \tag{A18}$$

The new parameters introduced in the above equilibrium conditions are

$$\lambda_I = (\lambda_q \lambda_z^{\alpha_2})^{\frac{1}{1-\alpha_1}},$$

$$\lambda_* = (\lambda_z^{\alpha_2} \lambda_q^{\alpha_1})^{\frac{1}{1-\alpha_1}},$$

$$\Delta \hat{\lambda}_t^* = \frac{1}{1-\alpha_1}(\alpha_1 \Delta \hat{q}_t + \alpha_2 \Delta \hat{z}_t),$$

$$\theta_p = \frac{\mu_p}{\mu_p - 1},$$

$$\kappa_p = \frac{(1-\beta\xi_p)(1-\xi_p)}{\xi_p},$$

$$\bar{\alpha} = \frac{1-\alpha_1-\alpha_2}{\alpha_1+\alpha_2},$$

$$\theta_w \equiv \frac{\mu_w}{\mu_w - 1},$$

$$\kappa_w = \frac{(1-\beta\xi_w)(1-\xi_w)}{\xi_w}.$$

Note that g_y is the average ratio of government spending to output, c_y is the average ratio of consumption to output, i_y is the average ratio of investment to output, μ_{pt} is the average price markup, μ_{wt} is the average wage markup, λ_q is the growth rate of investment-specific technology, λ_z is the growth rate of neutral technology, α_1 is the cost share of capital input, α_2 is the cost share of labor input, δ is the average capital depreciation rate, b is internal habit, S'' represents the investment adjustment costs, σ_u represents the curvature of the cost function of variable capital utilization, ξ_p is the probability that a firm cannot adjust its price, γ_p measures the degree of price indexation, ξ_w is a fraction of households who cannot reoptimize their wage decisions, and γ_w measures the degree of wage indexation.

In addition to all the equilibrium conditions, we have 7 shock processes:

$$\log \mu_{wt} = (1-\rho_w)\log \mu_w + \rho_w \log \mu_{w,t-1} + \sigma_w \varepsilon_{wt} - \phi_w \sigma_w \varepsilon_{w,t-1}, \text{ (price markup)}$$
$$\log \mu_{pt} = (1-\rho_p)\log \mu_p + \rho_p \log \mu_{p,t-1} + \sigma_p \varepsilon_{pt} - \phi_p \sigma_p \varepsilon_{p,t-1}, \text{ (wage markup)}$$
$$\log z_t = (1-\rho_z)\log z + \rho_z \log z_{t-1} + \sigma_z \varepsilon_{zt}, \text{ (neutral technology)}$$
$$\log q_t = (1-\rho_q)\log q + \rho_q \log q_{t-1} + \sigma_q \varepsilon_{qt}, \text{ (embodied technology)}$$
$$\log A_t = (1-\rho_a)\log A + \rho_a \log A_{t-1} + \sigma_a \varepsilon_{at}, \text{ (risk premium)}$$
$$\log \delta_t = (1-\rho_d)\log \delta + \rho_d \log \delta_{t-1} + \sigma_d \varepsilon_{dt}, \text{ (capital depreciation)}$$
$$\log \tilde{G}_t = (1-\rho_g)\log \tilde{G} + \rho_g \log \tilde{G}_{t-1} + \sigma_g \varepsilon_{gt} + \rho_{gz}\sigma_z \varepsilon_{zt}, \text{ (spending)}$$

where ε represents an i.i.d. normal shock and σ represents the corresponding standard deviation.

To compute the equilibrium, we eliminate both $\hat{u}_t$ and $\hat{r}_{kt}$ by using (A5) and (A8), leaving 9 equations and 9 variables $\hat{\pi}_t$, $\hat{w}_t$, $\hat{i}_t$, $\hat{q}_{kt}$, $\hat{c}_t$, $\hat{k}_t$, $\hat{y}_t$, $\hat{l}_t$, and $\hat{R}_t$. Out of these 9 variables, we have 7 corresponding observable variables (except $\hat{q}_{kt}$ and $\hat{k}_t$) for our estimation. Finally, we have one additional observable variable, the biased technology shock $\hat{q}_t$, used in our estimation.

In addition to the 9 equilibrium conditions, we have 7 equations describing the AR processes for the 7 structural shocks, 4 equations describing the 2 MA processes, and 7 equations concerning the 7 expectational terms in the system. Thus, there are 27 DSGE equations in total.

TABLE 1. Prior distributions of structural parameters

		Prior				
Parameters	Description	Distributions	α_{prior}	β_{prior}	5%	95%
General parameters						
b	Habit	Beta	1.0	2.0	0.025	0.776
α_1	Capital share	Beta	85.5869	159.4377	0.3	0.4
α_2	Labor share	Beta	38.4721	25.4535	0.5	0.7
η	1/(Frisch elasticity)	Gamma	1.0576	0.3106	0.2	10
$100(\lambda_q - 1)$	Biased tech growth	Gamma	1.8611	3.0112	0.1	1.5
$100(\lambda_* - 1)$	Output growth	Gamma	1.8611	3.0112	0.1	1.5
$100\,(\beta^{-1} - 1)$	Discount factor	Gamma	1.5832	1.0126	0.2	4.0
Firm parameters						
σ_u	Utilization cost	Gamma	3.7790	2.4791	0.5	3.0
S''	Adjustment cost	Gamma	1.0576	0.6213	0.5	5.0
$\mu_p - 1$	Price markup	Gamma	1.0	5.5	0.0094	0.5446
$\mu_w - 1$	Wage markup	Gamma	1.0	5.5	0.0094	0.5446
4δ	Depreciation	Beta	5.4257	41.4890	0.05	0.2
ξ_p	Calvo pricing	Beta	2.0384	3.0426	0.1	0.75
γ_p	Price indexation	Beta	1.0	1.0	0.05	0.95
ξ_w	Calvo wage	Beta	2.0384	3.0426	0.1	0.75
γ_w	Wage indexation	Beta	1.0	1.0	0.05	0.95
Policy parameters						
ρ_r	Interest persistence	Beta	1.0	2.0	0.025	0.776
ϕ_π	Inflation coef	Gamma	2.4373	1.0876	0.5	5.0
ϕ_y	Output coef	Gamma	1.0	1.0	0.05	3.0
$400 \log \pi^*$	Inflation target	Gamma	2.9043	0.7690	1.0	8.0

Note: "5%" and "95%" demarcate the low and high bounds of the 90% probability interval.

Table 2. Prior distributions of shock parameters

		Prior				
Parameters	Description	Distributions	α_{prior}	β_{prior}	5%	95%
Persistence parameters						
ρ_p	Price markup AR	Beta	1.0	2.0	0.025	0.776
ϕ_p	Price markup MA	Beta	1.0	2.0	0.025	0.776
ρ_w	Wage markup AR	Beta	1.0	2.0	0.025	0.776
ϕ_w	Wage markup MA	Beta	1.0	2.0	0.025	0.776
ρ_{gz}	Spending on tech	Gamma	1.8611	1.5056	0.2	3.0
ρ_a	Preference	Beta	1.0	2.0	0.025	0.776
ρ_q	Biased tech	Beta	1.0	1.0	0.05	0.95
ρ_z	Neutral tech	Beta	1.0	1.0	0.05	0.95
ρ_d	Depreciation	Beta	1.0	2.0	0.025	0.776
Standard deviations						
σ_r	Monetary policy	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_p	Price markup	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_w	Wage markup	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_g	Gov spending	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_z	Neutral tech	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_a	Preference	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_q	Biased tech	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_d	Depreciation	Inverse Gamma	0.4436	0.0009	0.0005	1.0
Transition matrix parameters						
q_{11}	DSGE model	Dirichlet	5.6667	1.0	0.5905	0.9911
q_{22}	BVAR model	Dirichlet	5.6667	1.0	0.5905	0.9911

Note: "5%" and "95%" demarcate the low and high bounds of the 90% probability interval.

TABLE 3. Posterior distributions of structural parameters

		DSGE model alone			Merged model		
Parameters	Description	Mode	5%	95%	Mode	5%	95%
General parameters							
b	Habit	0.544	0.493	0.624	0.528	0.597	0.954
α_1	Capital share	0.177	0.151	0.203	0.250	0.212	0.290
α_2	Labor share	0.804	0.747	0.818	0.679	0.614	0.740
η	1/(Frisch elasticity)	0.005	0.003	0.167	0.399	0.578	6.801
$100(\lambda_q - 1)$	Biased tech growth	1.507	1.215	1.911	1.438	1.145	1.700
$100(\lambda_* - 1)$	Output growth	0.483	0.400	0.569	0.519	0.221	0.576
$100\,(\beta^{-1} - 1)$	Discount factor	0.228	0.081	0.909	0.222	0.113	0.781
Firm parameters							
σ_u	Utilization cost	2.018	1.404	3.787	0.654	0.672	3.947
S''	Adjustment cost	0.800	0.608	1.278	0.710	0.495	3.032
$\mu_p - 1$	Price markup	0.000	0.000	0.001	0.000	0.000	0.017
$\mu_w - 1$	Wage markup	0.003	0.015	0.176	0.109	0.043	0.965
4δ	Depreciation	0.145	0.064	0.204	0.111	0.013	0.170
ξ_p	Calvo pricing	0.372	0.308	0.760	0.540	0.211	0.839
γ_p	Price indexation	0.121	0.028	0.408	0.394	0.024	0.721
ξ_w	Calvo wage	0.303	0.269	0.606	0.069	0.096	0.604
γ_w	Wage indexation	0.790	0.088	0.954	0.040	0.081	0.957
Policy parameters							
ρ_r	Interest persistence	0.618	0.572	0.687	0.477	0.490	0.744
ϕ_π	Inflation coef	1.480	1.392	1.693	2.008	1.820	3.230
ϕ_y	Output coef	0.066	0.052	0.101	0.141	0.099	0.228
$400\log\pi^*$	Inflation target	5.576	3.863	10.109	5.764	1.693	8.583

Note: "5%" and "95%" demarcate the low and high bounds of the 90% probability interval.

Table 4. Posterior distributions of shock parameters

		DSGE model alone			Merged model		
Parameters	Description	Mode	5%	95%	Mode	5%	95%
Persistence parameters							
ρ_p	Price markup AR	0.786	0.587	0.878	0.188	0.037	0.972
ϕ_p	Price markup MA	0.627	0.276	0.820	0.168	0.060	0.802
ρ_w	Wage markup AR	0.992	0.987	0.997	0.990	0.815	0.988
ϕ_w	Wage markup MA	0.530	0.305	0.827	0.000	0.040	0.695
ρ_{gz}	Spending on tech	0.947	0.490	1.348	1.690	0.490	2.138
ρ_a	Preference	0.988	0.973	0.995	0.988	0.242	0.953
ρ_q	Biased tech	0.994	0.988	0.997	0.989	0.971	0.993
ρ_z	Neutral tech	0.942	0.927	0.961	0.903	0.910	0.989
ρ_d	Depreciation	0.915	0.854	0.975	0.888	0.869	0.990
Standard deviations							
σ_r	Monetary policy	0.003	0.002	0.003	0.002	0.002	0.003
σ_p	Price markup	1.012	0.593	2.109	1.348	0.014	1.095
σ_w	Wage markup	0.023	0.017	0.065	0.009	0.016	0.404
σ_g	Gov spending	0.029	0.026	0.031	0.023	0.022	0.033
σ_z	Neutral tech	0.008	0.007	0.009	0.007	0.007	0.010
σ_a	Preference	0.061	0.035	0.137	0.030	0.013	0.075
σ_q	Biased tech	0.006	0.006	0.007	0.004	0.004	0.007
σ_d	Depreciation	0.096	0.065	0.261	0.098	0.069	1.008
Transition matrix parameters							
$q_{1,1}$	DSGE model				0.309	0.415	0.684
$q_{2,2}$	BVAR model				0.720	0.689	0.861

Note: "5%" and "95%" demarcate the low and high bounds of the 90% probability interval.

Table 5. Marginal data densities

Merged model	DSGE	BVAR
5848.90 - 5854.03	5735.49 - 5736.39	5685.74

Table 6. Output variance decompositions: contributions from a capital depreciation shock (%)

Quarters	4	8	12	16	20
Merged	48.60	47.89	43.77	40.54	38.58
DSGE alone	39.75	35.91	30.18	27.41	26.25

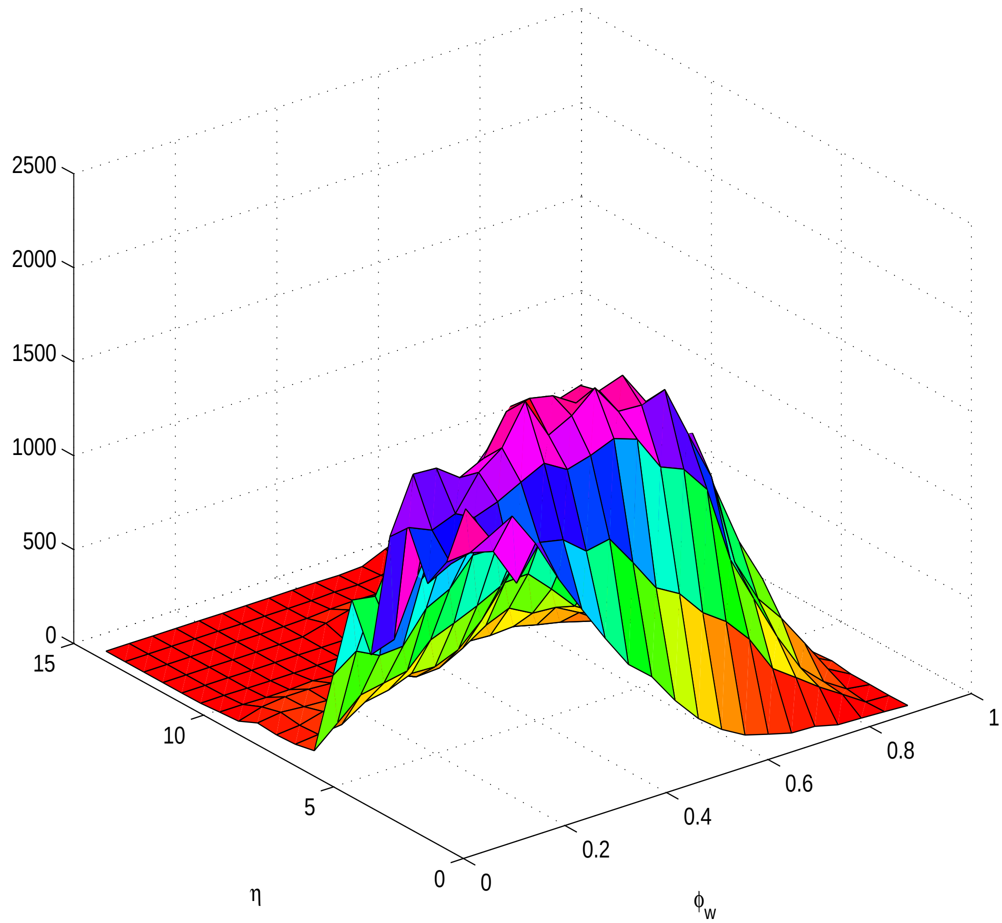

FIGURE 1. The joint posterior probability density of η and ϕ_w, after all the other parameters are integrated out through the posterior distribution. Note that η represents the inverse Frisch elasticity of labor supply and ϕ_w is the moving-average (MA) coefficient in the wage markup shock process.

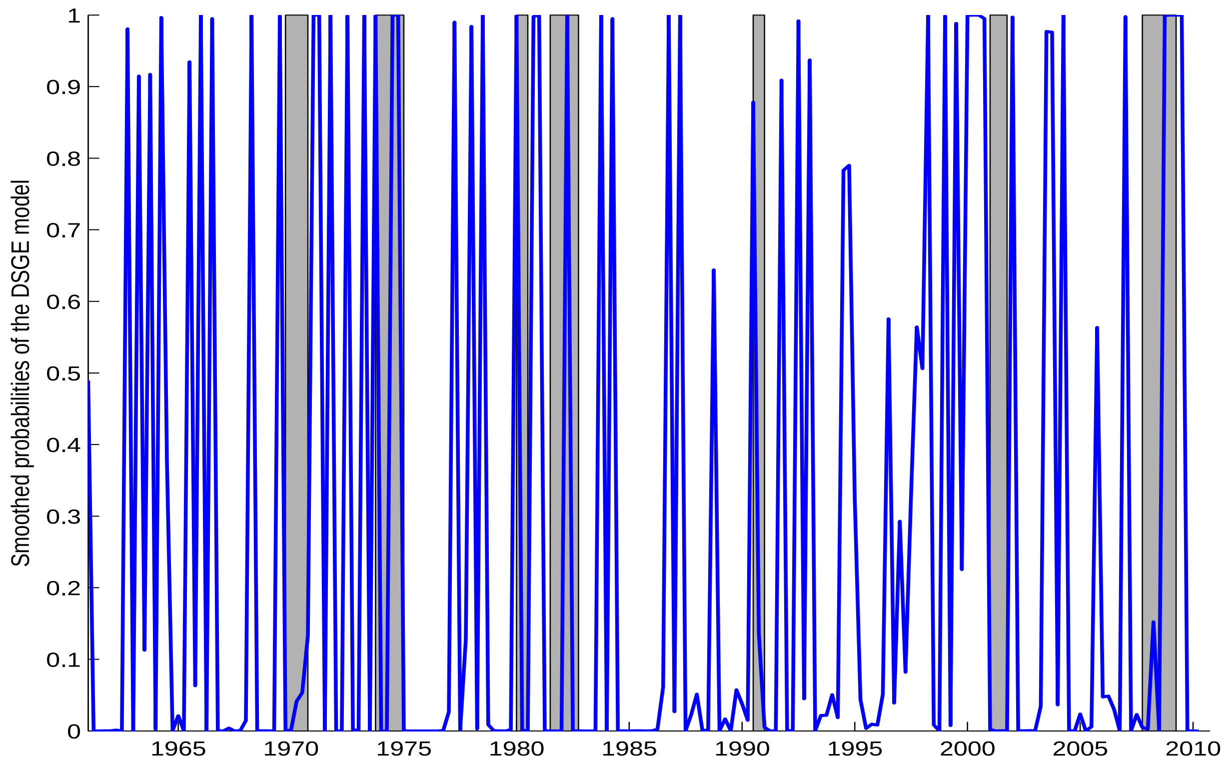

Figure 2. The posterior probabilities that the DSGE model is selected by the data. The shaded bars mark the NBER recession dates.

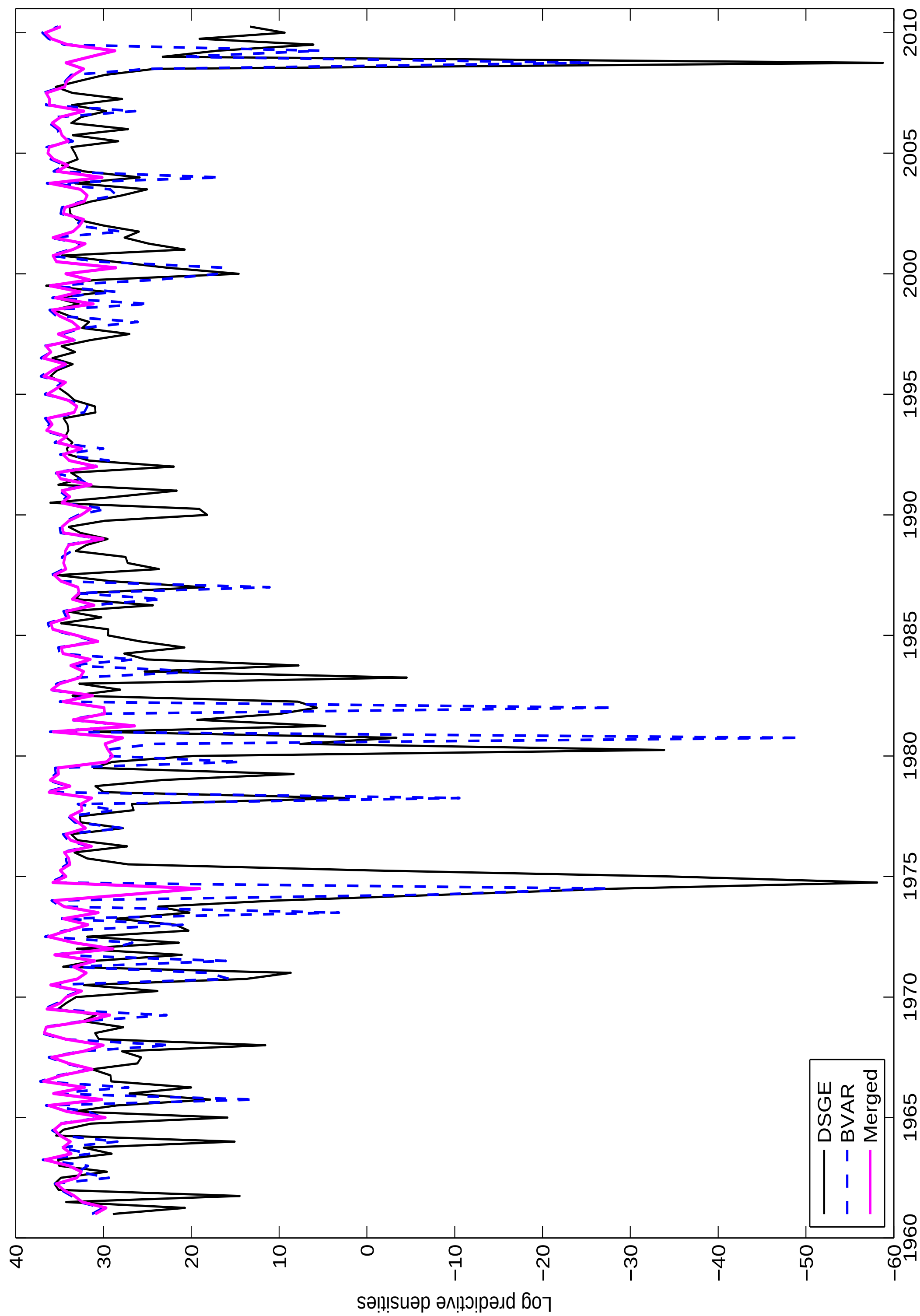

FIGURE 3. Log values of predictive densities from the three models.

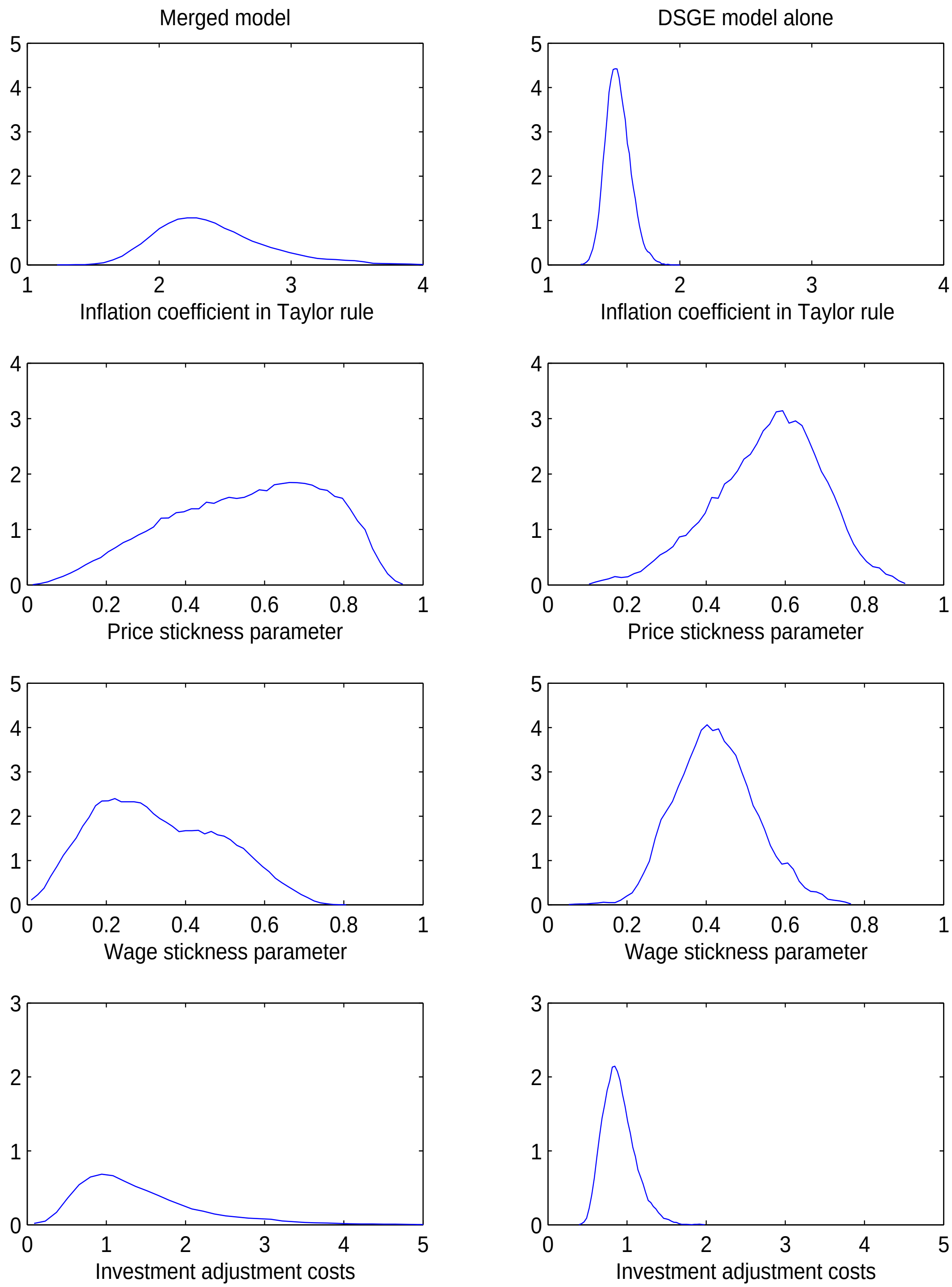

FIGURE 4. Marginal posterior distributions of some key structural parameters for the merged model (left column) and for the DSGE model when it is estimated in isolation (right column).

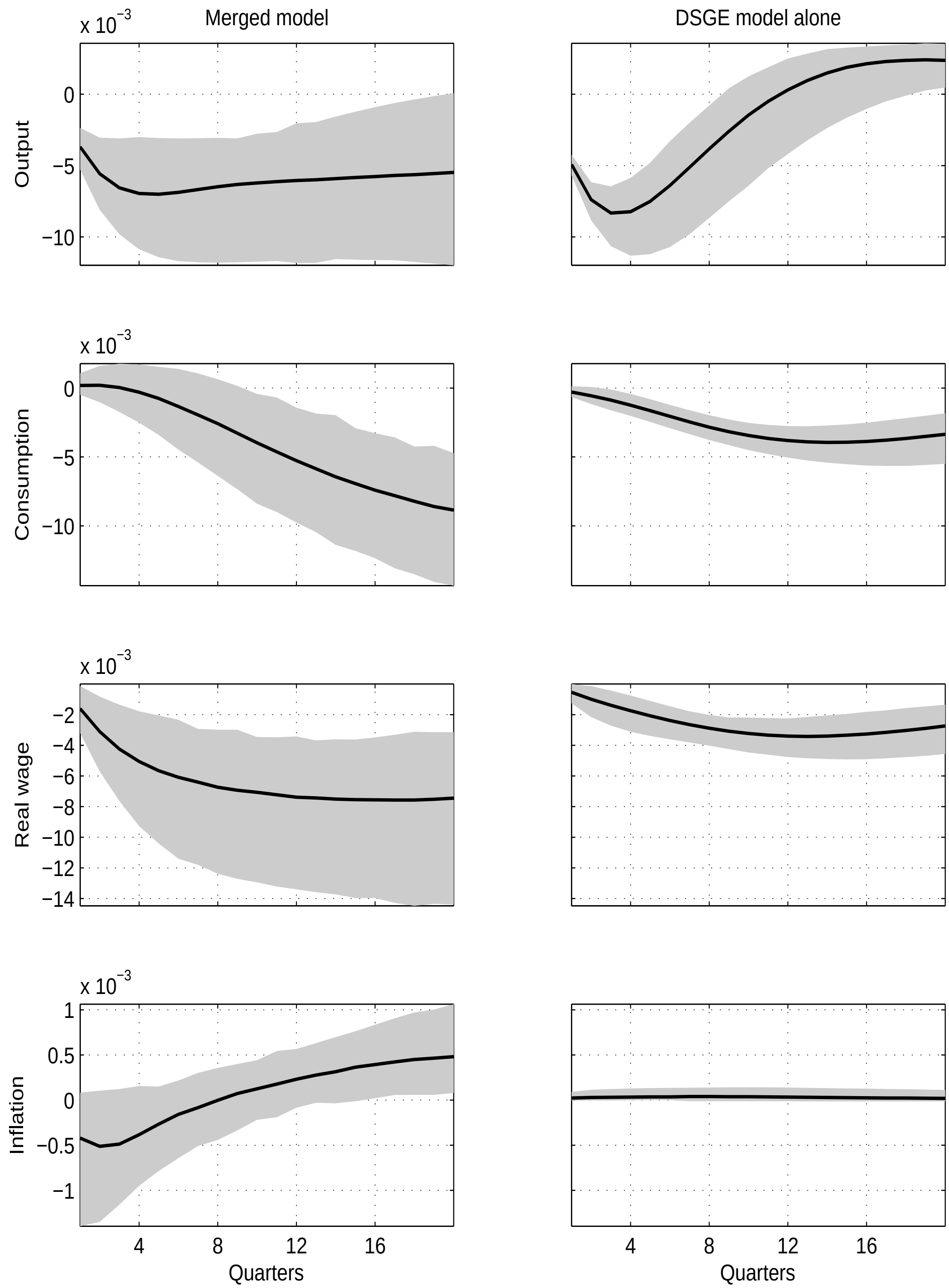

FIGURE 5. Impulse responses to a capital depreciation shock for the merged model (left column) and for the DSGE model when estimated in isolation (right column). The shaded area represents 90% posterior probability bands and the thick line represents the median estimate.

References

Altig, D., L. J. Christiano, M. Eichenbaum, and J. Linde (2004): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," Federal Reserve Bank of Cleveland Working Paper 04-16.

Brock, W. A., S. N. Durlauf, and K. D. West (2003): "Policy Evaluation in Uncertain Economic Environment," *Brookings Papers on Economic Activity*, 1, 235–301.

Chari, V., P. J. Kehoe, and E. R. McGrattan (2000): "Sticky Price Models of the Business Cycle: Can the Contract Multiplier Solve the Persistence Problem?," *Econometrica*, 68(5), 1151–1180.

Cogley, T., and T. J. Sargent (2005): "The Conquest of U.S. Inflation: Learning and Robustness to Model Uncertainty," *Review of Economic Dynamics*, 8, 528–563.

Cummins, J. G., and G. L. Violante (2002): "Investment-Specific Technical Change in the United States (1947-2000): Measurement and Macroeconomic Consequences," *Review of Economic Dynamics*, 5, 243–284.

Del Negro, M., and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45, 643–673.

——— (2008): "Forming Priors for DSGE Models (and How It Affects the Assessment of Nominal Rigidities)," Manuscript, Federal Reserve Bank of New York and University of Pennsylvania.

Denton, F. T. (1971): "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization," *Journal of the American Statistical Association*, 66, 99–102.

Diebold, F. X. (1991): "A Note on Bayesian Forecast Combination Procedures," in *Economic Structural Change: Analysis and Forecasting*, ed. by A. H. Westlund, and P. Hackl, pp. 225–232. Springer-Verlag, New York, NY.

Fisher, M., and D. F. Waggoner (2010): "Mixture Models and Bayesian Model Selection," Unpublished manuscript.

Geweke, J., and G. Amisano (forthcoming): "Optimal Prediction Pools," *Journal of Econometrics*.

Gordon, R. J. (1990): *The Measurement of Durable Goods Prices*. University of Chicago Press, Chicago,Illinois.

Hansen, L. P., and T. J. Sargent (2001): "Acknowledging Misspecification in Macroeconomic Theory," *Review of Economic Dynamics*, 4, 519–535.

Liu, Z., D. F. Waggoner, and T. Zha (2010): "Sources of Macroeconomic Fluctuations: A Regime-Switching DSGE Approach," Unpublished manuscript.

Sims, C. A. (2003): "Probability Models for Monetary Policy Decisions," Manuscript, Princeton University.

Sims, C. A., D. F. Waggoner, and T. Zha (2008): "Methods for Inference in Large Multiple-Equation Markov-Switching Models," *Journal of Econometrics*, 146(2), 255–274.

Sims, C. A., and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 949–968.

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

West, M., and J. Harrison (1997): *Bayesian Forecasting and Dynamic Models*. Springer, 2nd edn.

Federal Reserve Bank of Atlanta, Federal Reserve Bank of Atlanta and Emory University